2009-10

Understanding Uncertainty in Climate Model Components

Tokmakian, Robin; Challenor, P.; Gattiker, Jim

Tokmakian, R., Understanding Uncertainty in Climate Model Components, NOAA's 34th Climate Diagnostics and Prediction Workshop, October 26-30, 2009 Monterey, CA
http://hdl.handle.net/10945/43800

Understanding Uncertainty in Climate Model Components

Robin Tokmakian
Naval Postgraduate School rtt@nps.edu

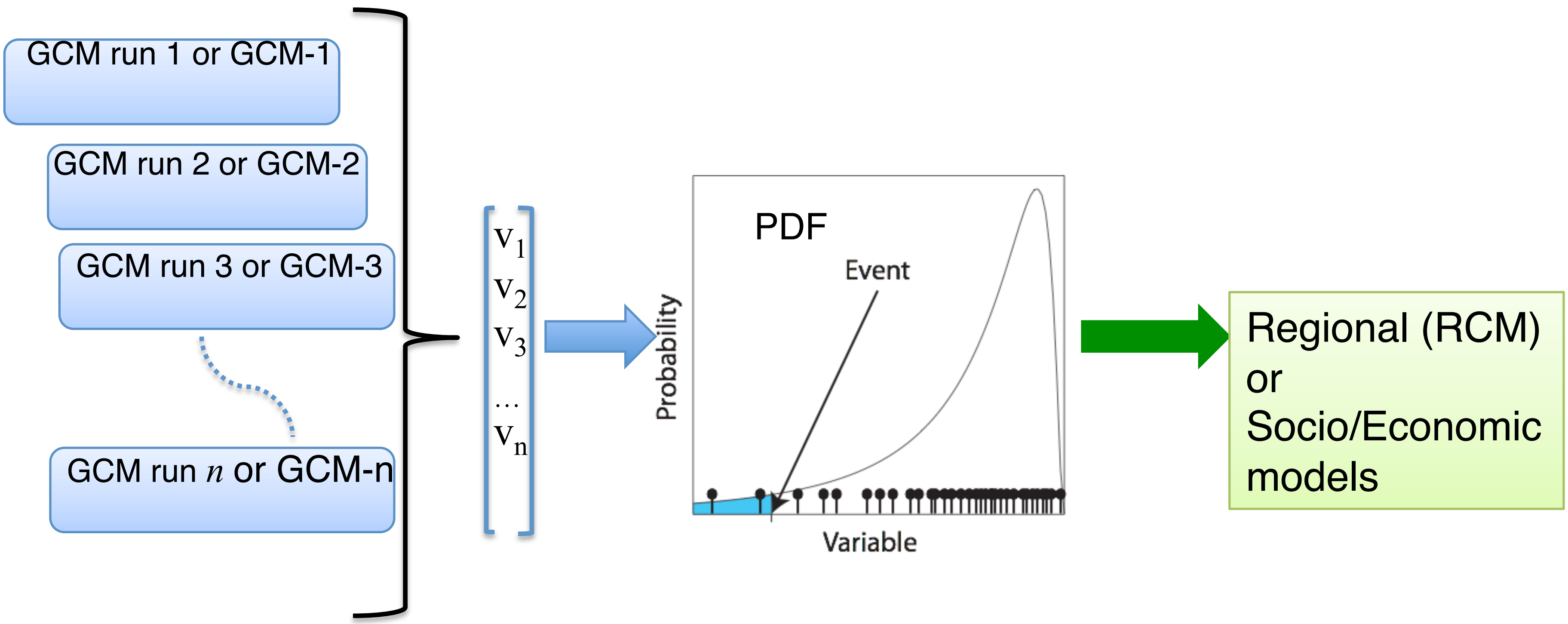
Collaborators: P. Challenor National Oceanography Centre, UK;
Jim Gattiker Los Alamos National Laboratory

http://www.oc.nps.edu/~rtt/Pages/ASSURE.html

- Introduction & Motivation
- Uncertainty Methodology
- Designing the Experiment
- Ocean Metrics
- Early Outcomes

NOAA's 34th Climate Diagnostics and Prediction Workshop
October 26-30, 2009 Monterey, CA

Uncertainty and flow of information

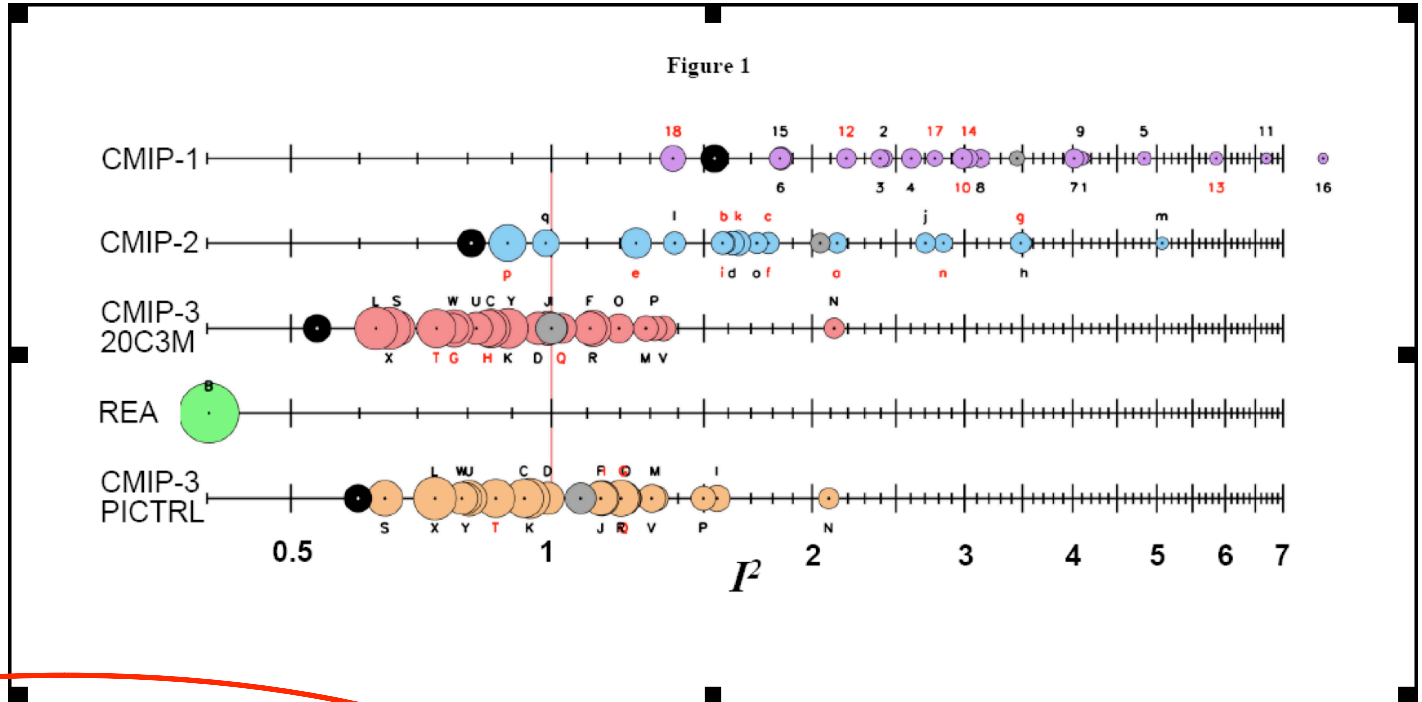


NOAA's 34th Climate Diagnostics and Prediction Workshop
October 26-30, 2009 Monterey, CA

Performance index *I2.*

- Circles sizes: 95% C.I.
- Grey: average within one model group.
- Black circles: multi-model mean
- Green circle NCEP REA

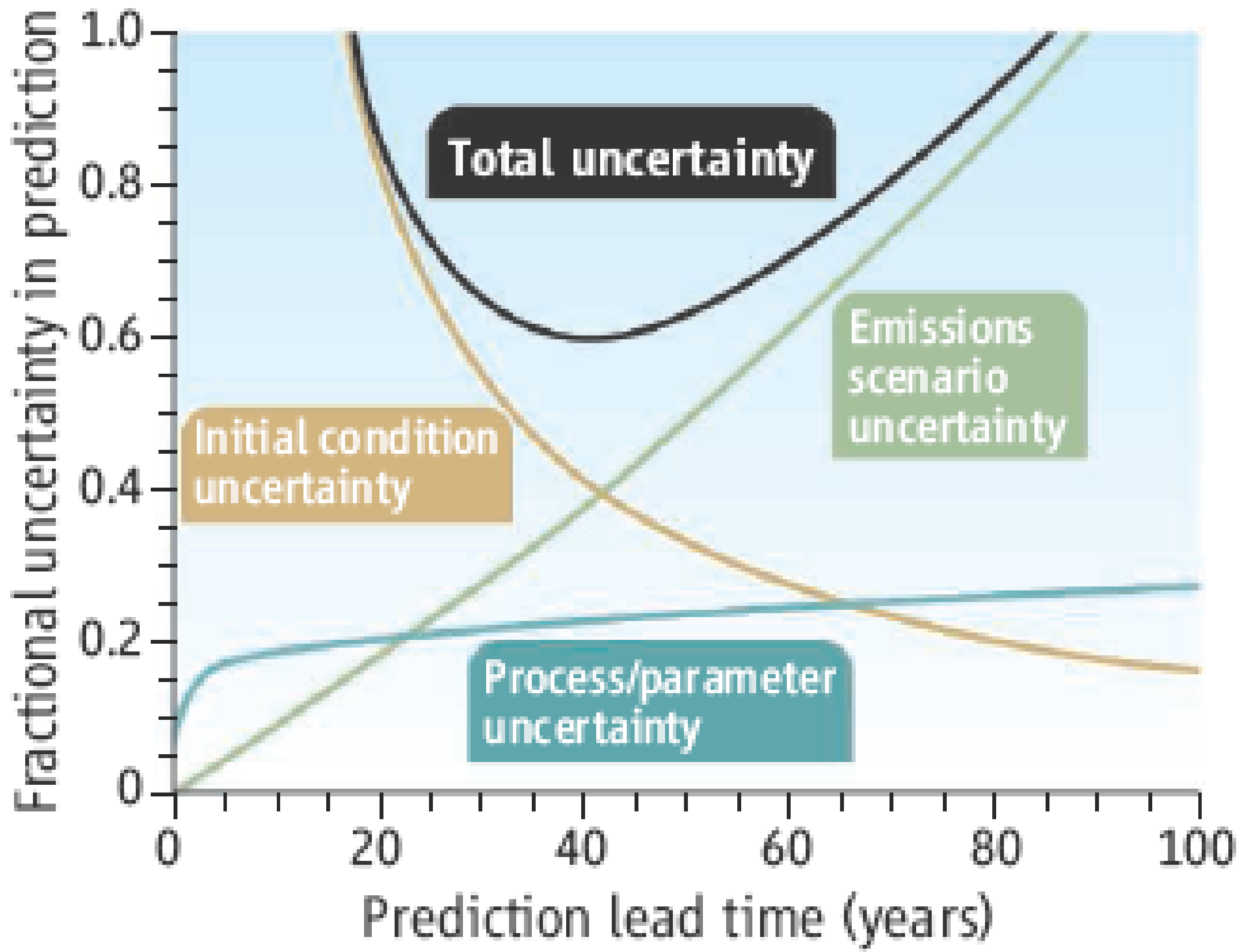
(from Reichler and Kim, 2008) Bader et al. CCSP 3.1 2009



$$I_{vn}^2 = \frac{\sum_n \left(w_n (\overline{s_{vn}} - \overline{o_{vn}})^2 / \sigma_{vn}^2 \right)}{\sum_n \left(w_n (\overline{s_{vn}} - \overline{o_{vn}})^2 / \sigma_{vn}^2 \right)^{m=20C3M}}$$

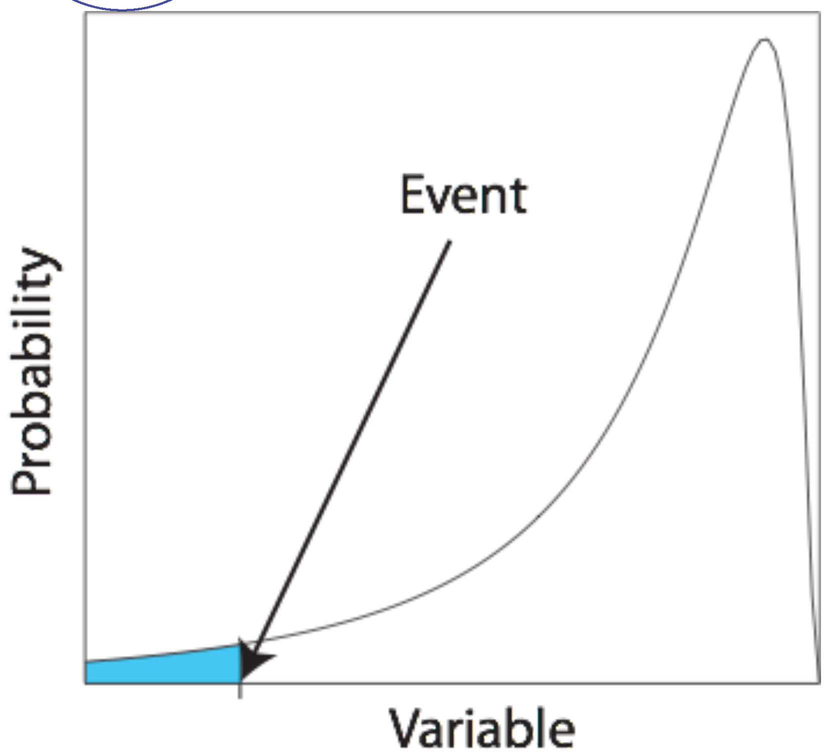
s_{vn} : climatology for climate variable (v) model (m), and grid point (n)
o_{vn}: observed climatology
w_n : weights needed for area and mass avg.
σ^2_{Vn} : interannual variance from the obs.

$$I_m^2 = \overline{I_{vn}^2}$$



Cox and Stephenson, Science 2007

Uncertainty estimation requires a PDF be created that represents the sampling of the full parameter space.



We want to estimate the area of the proportion of model predictions to the left of the tipping point

Question: How do we create a *single model ensemble* that will be representative of a full PDF ?

Method 1: Monte Carlo Methods → Very Large Complex Model Ensemble

GCM run 1
With parameter array X_1: $[x_1, x_2, x_3 \ldots x_{10}]$

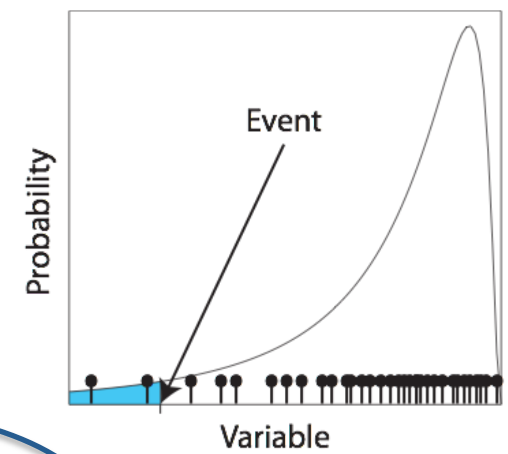
GCM run 2
With parameter array X_2: $[x_1, x_2, x_3 \ldots x_{10}]$

GCM run 3
With parameter array X_3: $[x_1, x_2, x_3 \ldots x_{10}]$

GCM run n
With parameter array X_n: $[x_1, x_2, x_3 \ldots x_{10}]$

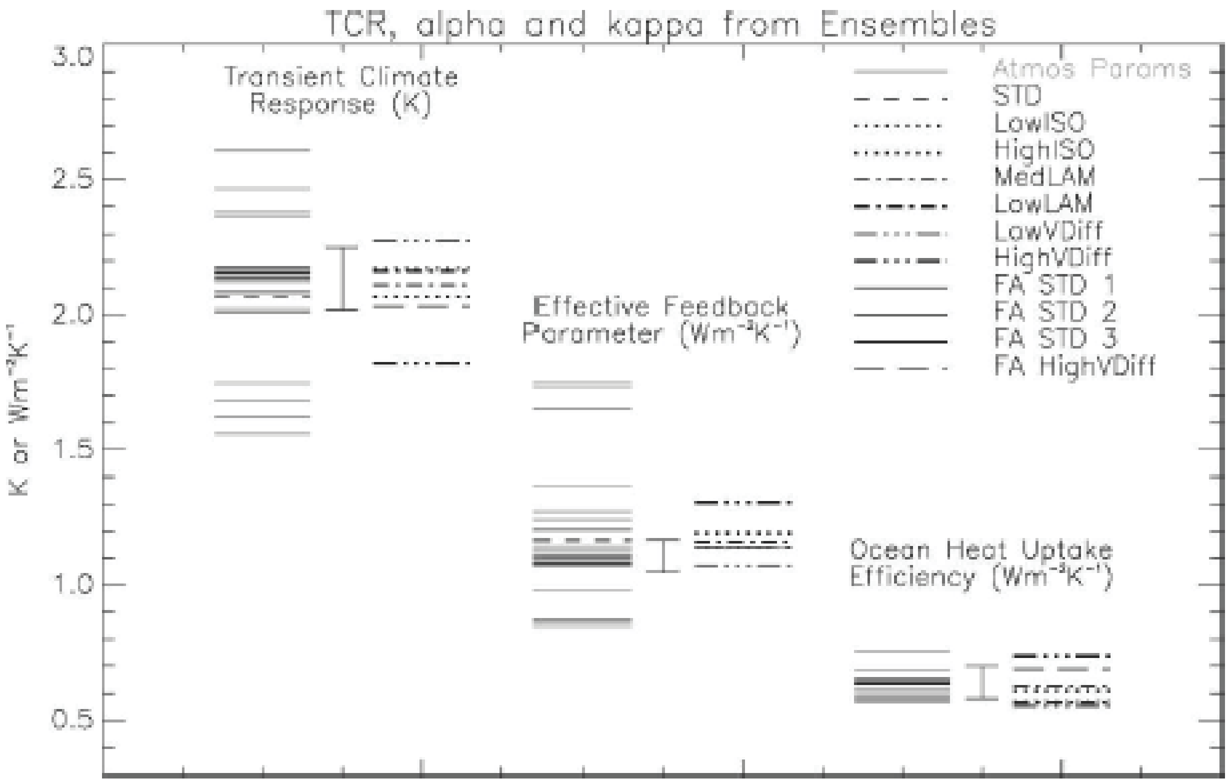
$$\begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ \ldots \\ V_n \end{bmatrix}$$

Full PDF



n = O(10,000)

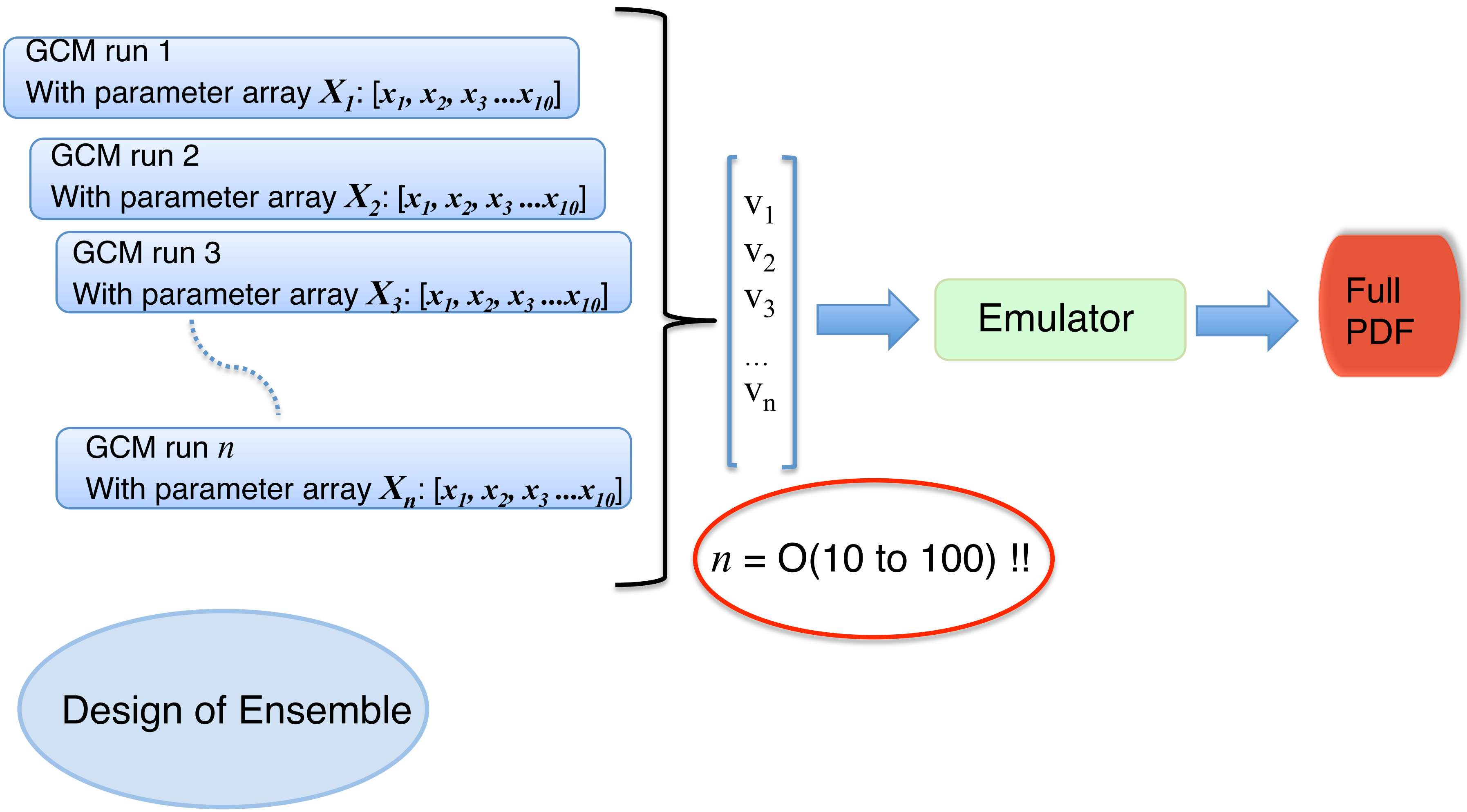
Single Model Ensembles



Collins et al 2007; Perturbed ensemble HADCM3 (16 members)

Climateprediction.net – 10,000+ simulations; async. communications between ocean/atm HADCM3

Method 2 – Complex model + PDF emulator



A Designed Experiment

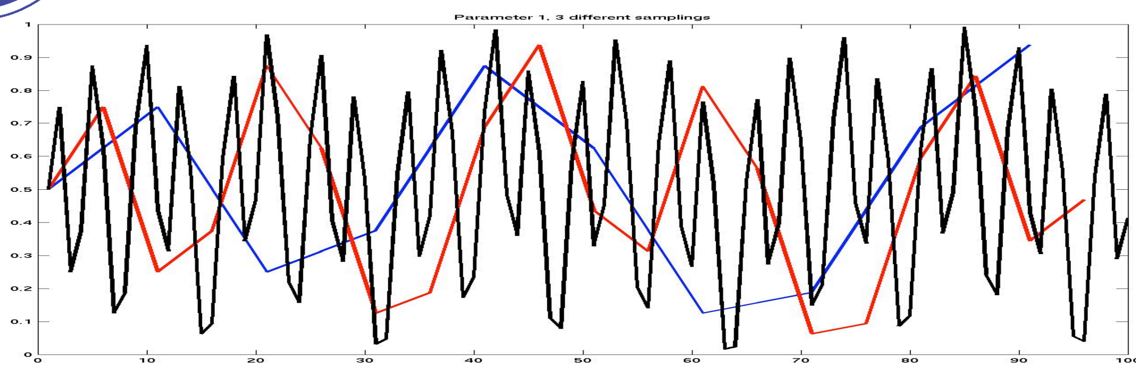
Goal: Create a set of parameters that sample the model space adequately

- Method 1: Create a complete set of parameter permutations, with adequate incremental sampling.
    - *With 10 parameters, permutations of GCM simulations necessary to run will expand exponentially*.

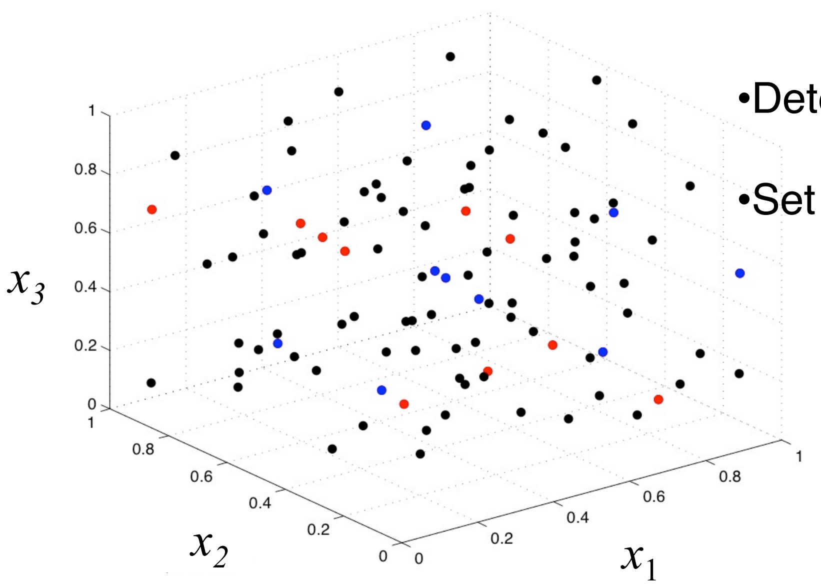
- Method 2: Use Sobol Sequence or Latin Hypercube methods to span parameter space to reduce the number of simulations required.

[Santner TJ, Williams B, Notz W., 2003, The design and analysis of computer experiments. New York: Springer]

Sobol Sequences



Blue 10 runs
Red 20 runs
Black 100 runs

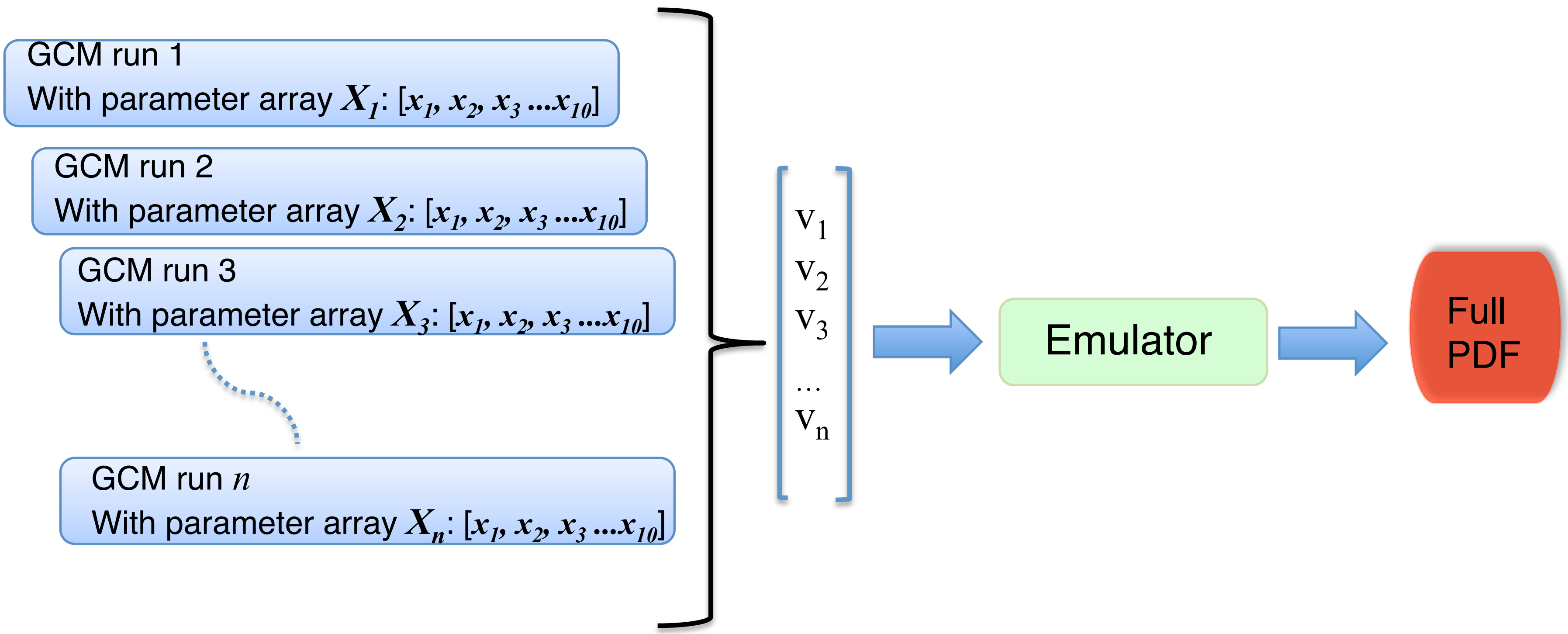


•Determine min/max values of each parameter

•Set appropriately using scale from 0 →1

e.g. Press, W. et al. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, Second Edition, 1992; Sobol,I.M., Distribution of points in a cube and approximate evaluation of integrals, *Comput. Maths. Math. Phys.,*1967

Method 2 – Complex model + PDF emulator



Design of Ensemble

Emulator Details

Emulator Details

DACE (Design and Analysis of Computer Experiments) or
BACCO (Bayesian Analysis of Computer Code Output).

Given a climate model: $Y = F(x)$; with vector x as tunable inputs

Using a "small" set of simulations or runs, varying the values of x

Build an emulator, $f(x)$ for $F(x)$ with the following characteristics:

- reflects the true value of Y at points x
- at other points, the distribution of $F(x)$ should give a mean value for $F(x)$ that represents a plausible value of Y given any vector x
- the probability distribution should be a realistic view of the uncertainty in the approximation to the full model.

Emulator has 2 parts:

- a mean function
- a zero-mean Gaussian process representing the non-linearities in $F(x)$

Mean process

$$m_0(x) = h(x)^T b$$

q coefficient vector

vector of *q* regression functions

Gaussian process

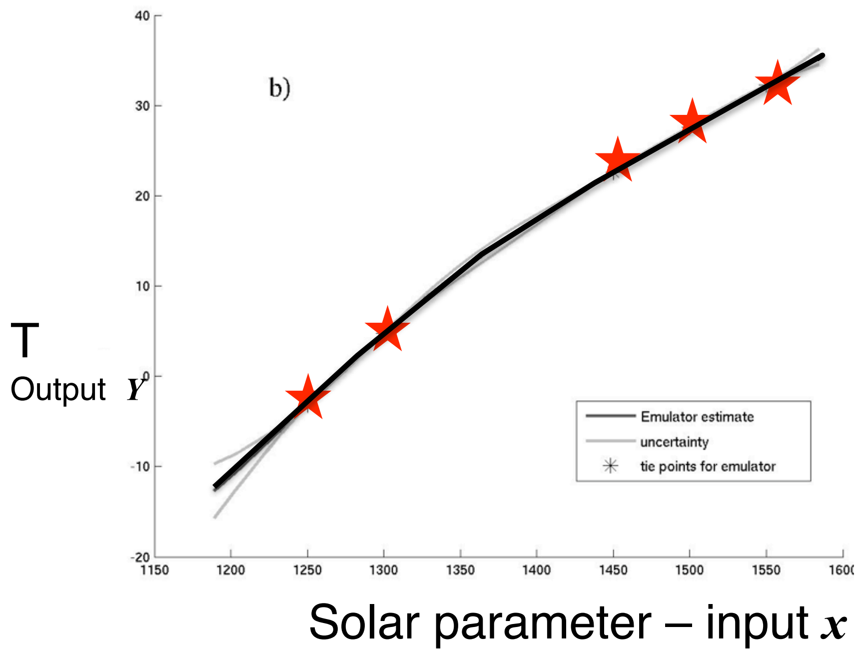
$$GP(x) = \sigma^2 \underbrace{e^{(-(x_1 - x_2)^T B (x_1 - x_2))}}$$

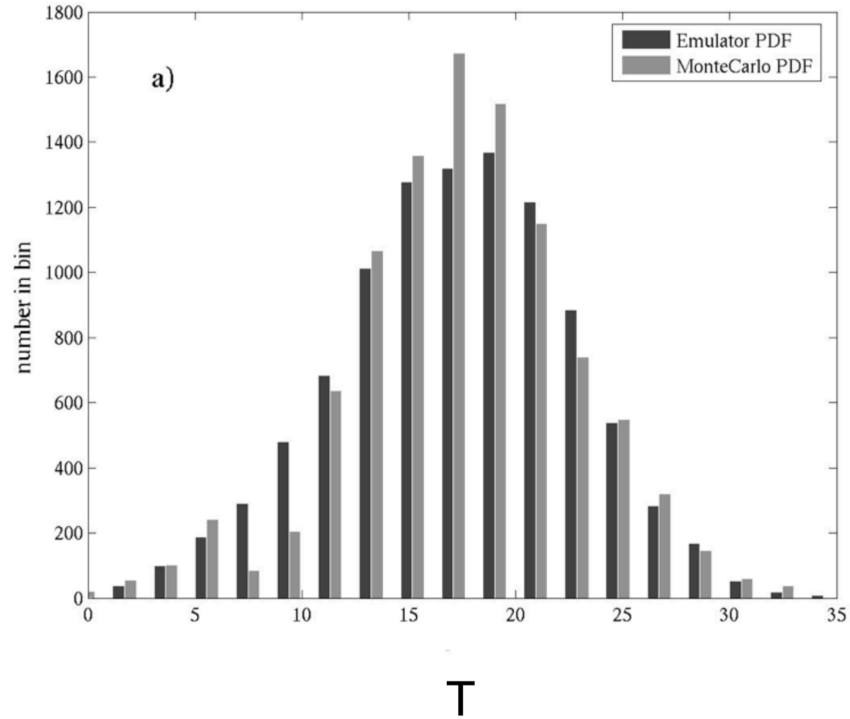
Variance

A Gaussian process which assumes stationarity

$$F(x) \sim f(x) = m_0(x) + GP(x)$$

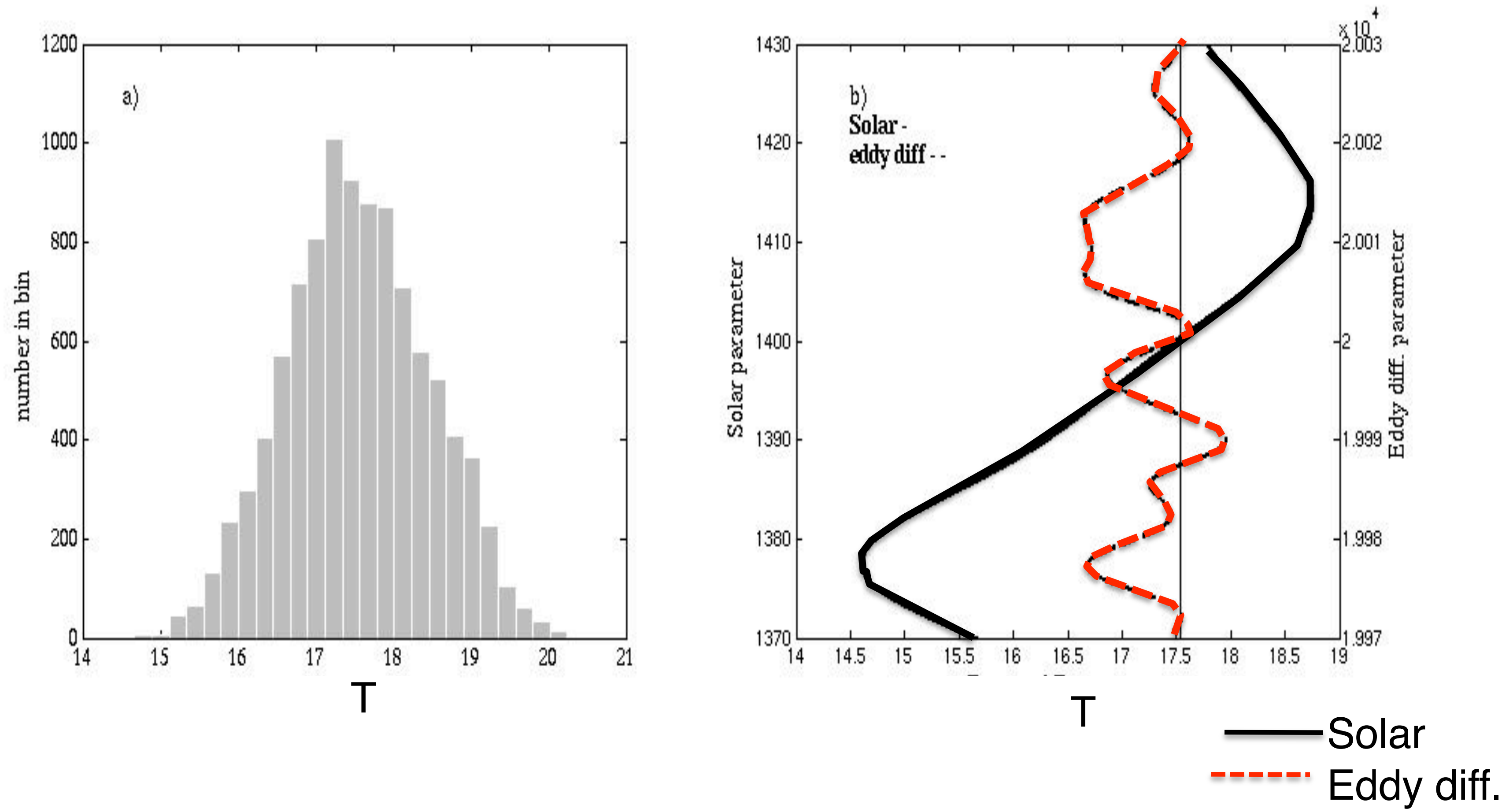
A Simple Example





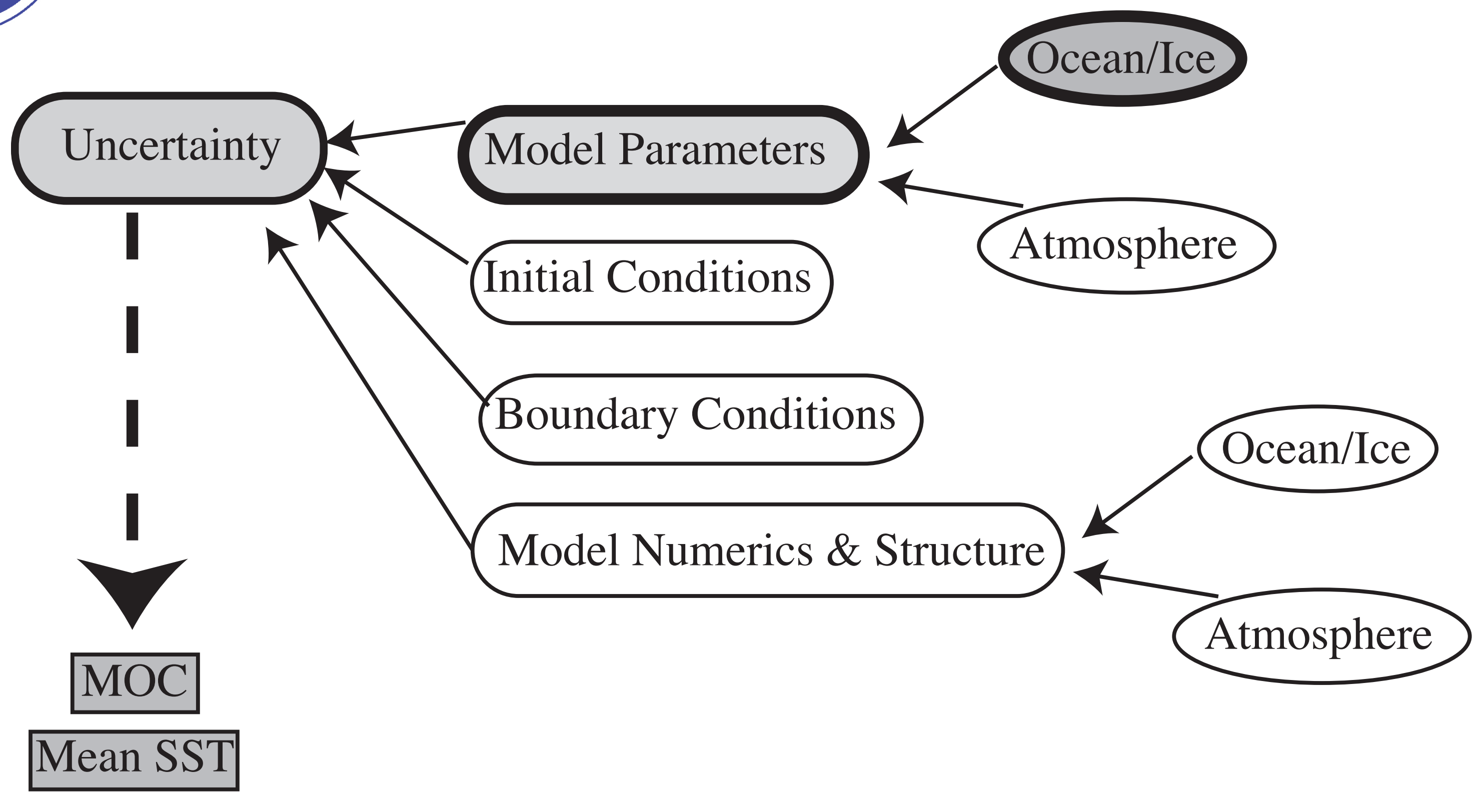
- Using a simple energy balance model;
- Run 5 times, with varying values of a solar related parameter
- Resulting PDF is in Black on right side.
- Compare to PDF created from Monte Carlo method (grey)

An Example Using a vector x with 2 parameters



NOAA's 34th Climate Diagnostics and Prediction Workshop
October 26-30, 2009 Monterey, CA

A feasibility study using a GCM



A feasibility study using a GCM

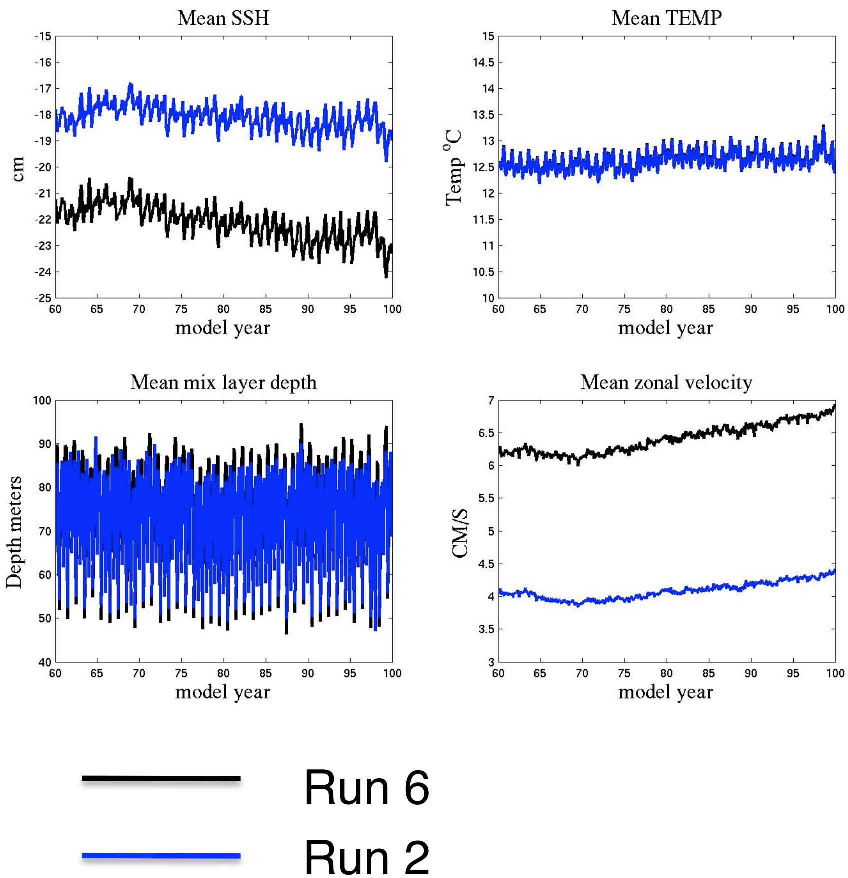
- CCSM3.0 x 3 (~ 3 degree resolution)
- Active ocean/ice components (POP2 & CICE)
- NCEP inter-annual reanalyzes forcing
- 100 member ensemble
- 9 parameters
- Initial Design Phase (10 runs)
- Full experiment phase (100 runs)

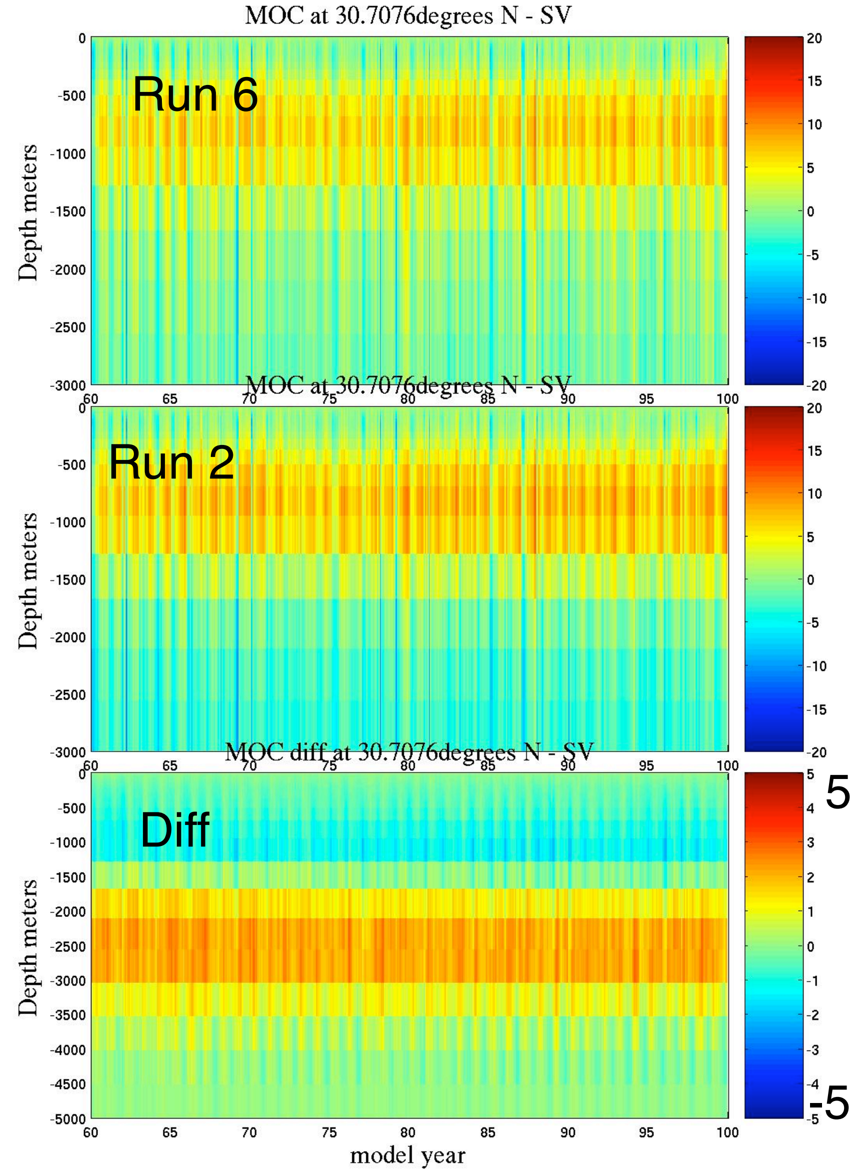
A feasibility study using a GCM

RUN #	b_vdc1	b_vdc_depth	ah_gm/bolus	slm	vconst_1/6	convt_diff	convt_visc	albicev +	albsnowv +
1	2.53	1.50E+05	3.05E+07	0.155	5.05E+07	50000	50000	0.60	0.90
2	3.76	1.00E+05	4.53E+07	0.083	7.53E+07	25001	75000	0.50	0.93
3	1.29	2.00E+05	1.58E+07	0.228	2.58E+07	75000	25001	0.70	0.88
4	1.91	1.25E+05	3.79E+07	0.046	8.76E+07	87500	12501	0.65	0.94
5	4.38	2.25E+05	8.38E+06	0.191	3.81E+07	37501	62500	0.45	0.89
6	3.14	0.75E+05	2.31E+07	0.119	1.34E+07	62500	87500	0.75	0.92
7	0.67	1.75E+05	5.26E+07	0.264	6.29E+07	12501	37501	0.55	0.86
8	0.98	1.13E+05	1.94E+07	0.209	5.67E+07	18751	6251	0.78	0.93
9	3.45	2.13E+05	4.89E+07	0.064	7.19E+06	68750	56250	0.58	0.88
10	4.69	0.63E+05	3.42E+07	0.282	3.19E+07	43751	81250	0.68	0.91

Examples of metrics to be examined

- SST – regional and global; mean and variance
- Mixed layer depths
- Transports – Heat, volume; across basins, passages
- Heat Content
- Current strengths, locations
- Meridional overturning strength





NOAA's 34th Climate Diagnostics and Prediction Workshop
October 26-30, 2009 Monterey, CA

Long term goals

- Relate, through statistical methods, low resolution to higher resolution models

- Apply methods to full CCSM to include atmospheric parameters evaluation

- Use methodology to examine uncertainty in initial conditions

    - Much shorter GCM runs → allows for higher resolution models to be used

    - Includes investigating methods to reduce size of initial condition space such as EOFs.

Thank you for your attention

References:

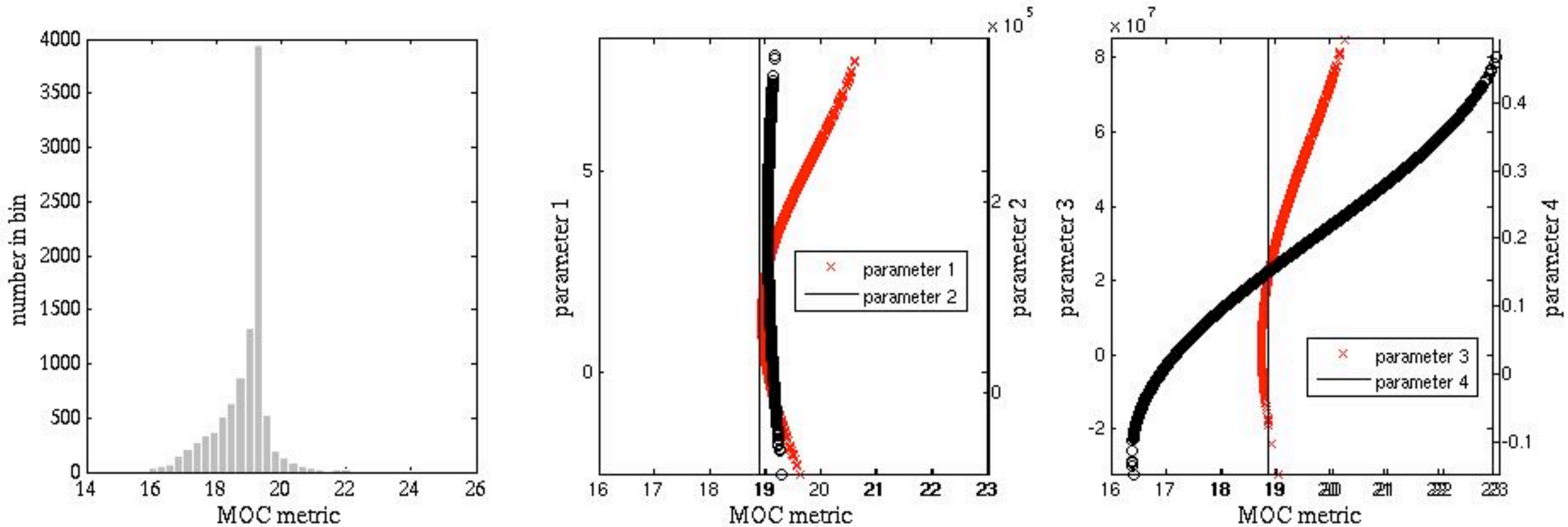
•Challenor, P.G, McNeall, D, Gattiker, J., 2009 Estimating the Probability of Rare Climate Events in The Handbook of Applied Bayesian Analysis, Editor A O.Hagan and M. West , Oxford University Press, in press.

•Gattiker,J., 2005, Using the Gaussian Process Model for Simulation Analysis Code, Los Alamos technical report LA-UR-05-5215.

•Higdon, D., J. Gattiker, B. Williams, and M. Rightley, 2008, Computer Model Calibration Using High-Dimensional Output, J. of the American Statistical Association, 103,482, Applications and Case Studies, DOI 10.1198/016214507000000888

•O'Hagan, A., 2006, Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering and System Safety,* **91**, 1290–1300.

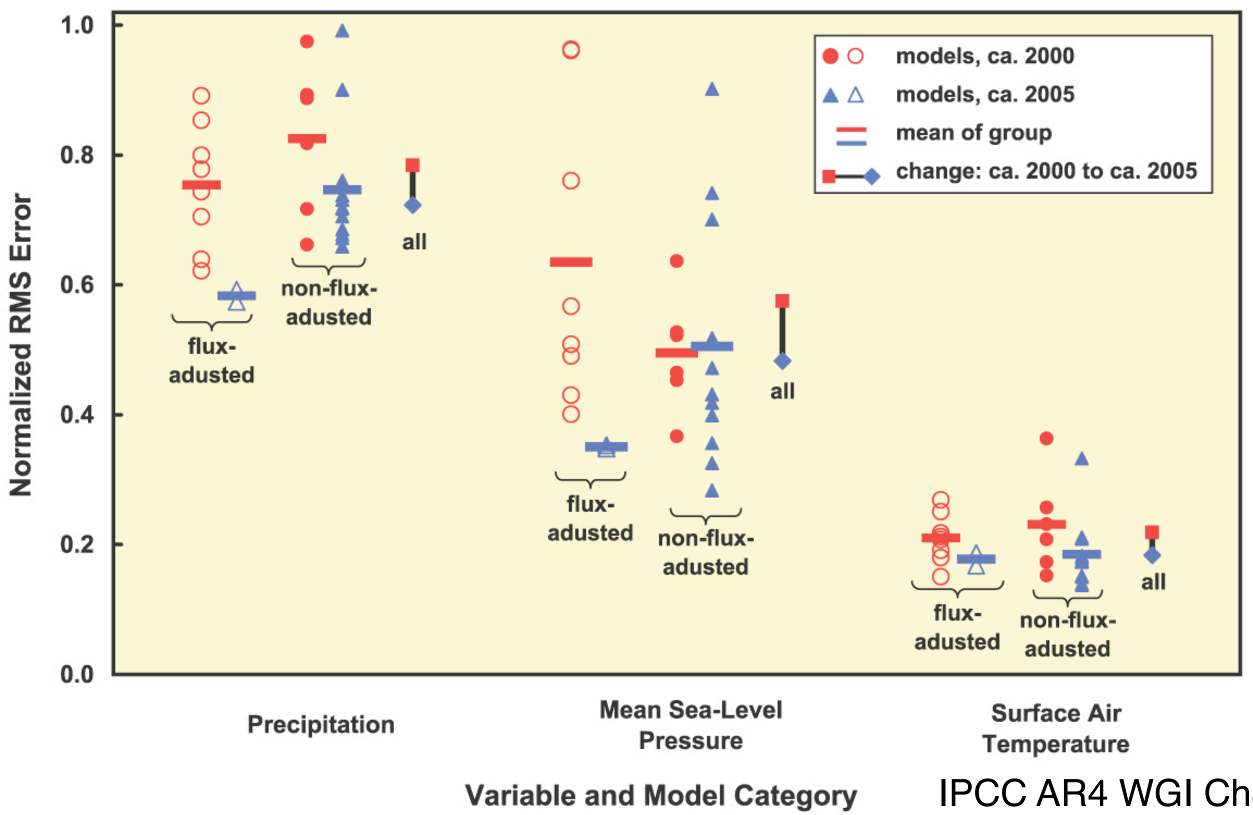
Estimating the uncertainty with 4 parameters and 10 estimates of MOC



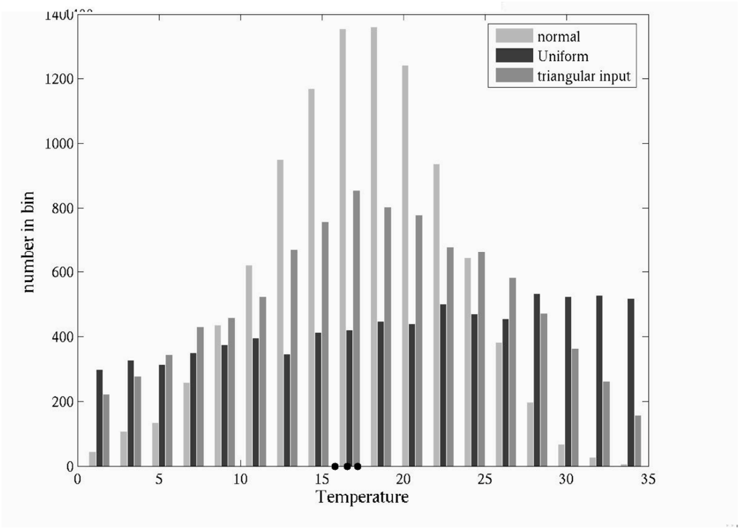
In this “test” example:

- All parameters make some difference in resulting MOC value
- Parameter 2 has little impact
- The output is very sensitive to the setting of parameter 4

Multi-model Uncertainty Methods



IPCC AR4 WGI Ch8 Fig.8-11



normal
Uniform
triangular input
number in bin
Temperature
0
200
400
600
800
1000
1200
1400
0
5
10
15
20
25
30
35